Static and dynamic cues in vowel production in Hijazi Arabic

Wael Almurashi,[1] Jalal Al-Tamimi,[1] and Ghada Khattab[1]

[1]*Speech and Language Sciences, School of Education, Communication and Language Sciences, Newcastle University, King George VI Building, Newcastle upon Tyne NE1 7RU, UK*

Static cues such as formant measurements obtained at the vowel midpoint are usually taken as the main correlate for vowel identification. However, dynamic cues such as vowel-inherent spectral change (VISC) have been shown to yield better classification of vowels using discriminant analysis. The aim of this study is to evaluate the role of static versus dynamic cues in Hijazi Arabic (HA) vowel classification, in addition to vowel duration and F3, which are not usually looked at. Data from 12 male HA speakers producing eight HA vowels in /hVd/ syllables were obtained, and classification accuracy was evaluated using discriminant analysis. Dynamic cues, particularly the three-point model, had higher classification rates (average 95.5%) than the remaining models (static model: 93.5%; other dynamic models: between 65.75% and 94.25%). Vowel duration had a significant role in classification accuracy (average +8%). These results are in line with dynamic approaches to vowel classification and highlight the relative importance of cues such as vowel duration across languages, particularly where it is prominent in the phonology.

I. INTRODUCTION

Formant frequencies are crucial acoustic correlates for the identification of vowels. For many years, however, the main approach to describing vowels has focused on measuring the first two formants (F1 and F2) at steady-state (e.g., Peterson and Barney, 1952). This static approach is extensively followed because it is believed that measuring a single sample of the monophthong vowels (e.g., midpoint), where shifts in formant values are typically minimal, yields the target position a speaker tries to reach when he/she

produces vowels (e.g., Munro, 1993; Almbark and Hellmuth, 2015). Additionally, it reflects the vowel target from an articulatory, acoustic and perceptual point of view (Strange, 1989).

Nevertheless, subsequent studies, primarily from varieties of English, have reported that reducing vowels' acoustic portrayal to a static parameterization has important limitations and have noted other cues such as dynamic cues that can define vowel characteristics effectively. To illustrate, dynamic cues—in particular, vowel-inherent spectral changes (VISC) (e.g., Nearey and Assmann, 1986; Huang, 1992; Harrington and Cassidy, 1994; Hillenbrand et al., 1995; Arnaud et al., 2011; Morrison and Assmann, 2012)—contain essential information (e.g., spectral movements), not only for diphthong vowels but also for monophthong vowels, which cannot be represented adequately by taking a single point from the vowel. Moreover, investigating vowel discrimination using a perception test with "silent center" vowels (e.g., Strange, 1989) revealed that the spectral information around the vowel's onset and offset has a bigger influence on identifying vowels than centers do. Additionally, accounting for VISC can yield better identification of monophthongs when the acoustic parameters are taken from more than one location by using discriminant analysis, which is a statistical method that many studies have used in predicting listeners' categorization patterns (e.g., Hillenbrand et al., 1995; Hillenbrand et al., 2001; Arnaud et al., 2011). More specifically, discriminant analysis classifies items (e.g., vowels) into discrete categories using acoustic measures as input parameters and then shows the percentages of how well the vowels could be separated based on their acoustic measurements.

A VISC is defined by Nearey and Assmann (1986) as the "relatively slowly varying changes in formant frequencies associated with vowels themselves." It is based on the assumption that the formant trajectories of the studied vowels can be specified by shifts in frequencies when measurements are taken from more than one location between the vowel's onset (at around 20%) and the vowel's offset (at around 80%) over the full duration of the vowel. The VISC approach aims to evaluate inherent vowel variation (e.g., the

slowly varying shifts in the vowel formant values) alongside the vowel target after eliminating the effects of surrounding consonants, and it is usually described as intrinsically dynamic (Nearey and Assmann, 1986; Hillenbrand et al., 1995). There are three primary accounts of VISC with competing acoustic parameterizations (Nearey and Assmann, 1986; Gottfried et al., 1993; Morrison and Nearey, 2007; Arnaud et al., 2011). All three approaches highlight the relevance of a sample formant pattern taken around the onset, but they do not agree on which additional cues are significant. The first approach is onset + offset (offset model, henceforth), in which the formant frequencies of the onset and also of the offset are what matters. The second approach is onset + slope (slope model, henceforth), which argues that the rate of change over time is the significant cue. The third approach is onset + direction (direction model, henceforth), which states that what is important is the general direction of formant frequency changes.

Many studies, mostly from English varieties/languages (e.g., Nearey and Assmann, 1986; Hillenbrand and colleagues, 1995; 1999; 2001; Morrison and Assmann, 2012), have compared static spectral features with one or all of the approaches of VISC to discover the extent to which formant frequency shifts can contribute to the separation of vowel categories. They found that including the VISC parameterization outperforms all approaches based on static spectral characteristics. For example, Hillenbrand and colleagues (1995; 1999; 2001) concluded that using two points—namely, onset (around 20%) + offset (around 80%) parameterizations of American English diphthongs and monophthongs—leads to higher classification accuracies than using one point located nearer the steady-state of the vowel. Others found that the three-point model (where formant measurements are taken from three locations, namely, at 20% onset, 50% midpoint, and 80% offset during vowel duration) yields more accurate vowel separation than the midpoint model (static approach) (e.g., Huang, 1992; Zahorian and Jagharghi, 1993; Harrington and Cassidy, 1994; Hillenbrand et al., 1995; Ferguson and Kewley-Port, 2002). For example, Huang (1992) concluded that classification accuracy increases when triple samples are taken from the vowel compared to using one sample. In a similar vein, Hillenbrand et al. (1995) reported that taking three measurements

outperformed the one-point model but provides little improvement over the offset model. Some studies (e.g., Arnaud et al., 2011, on Canadian French) that tested all three VISC models against the static model have even concluded that classification accuracy increases when using VISC approaches compared to using just one sample, no matter the combination of acoustic parameters.

Another line of studies (e.g., Watson and Harrington, 1999; Slifka, 2003) found VISC models helpful in improving the separation between lax and tense vowels. For example, Slifka (2003) found that using the slope was useful in the classifications of tense/lax vowels in English, with the lax vowels having a rising slope (positive) and the tense vowels having a falling slope (negative). Others have evaluated the suitability of a VISC approach as a function of the density of a particular vowel system. For instance, using the offset approach, Jin and Liu (2013) researched the degree of spectral shift of five vowels, namely, /ɑ, o, e, i, u/, spoken by Chinese (CN) speakers and Korean (KN) speakers whose vowels were recorded in the context of /hVda/, following the phonological structures of Mandarin CN and KN. They found that CN speakers, who have a sparse vowel system (six monophthongs), displayed significantly greater spectral shifts of vowels than KN speakers, who have a dense vowel system (ten monophthongs). This corroborated the findings of Manuel (1990), Meunier et al. (2003), and Al-Tamimi and Ferragne (2005). On the other hand, an opposing view is held by other researchers (e.g., Hillenbrand, 2013; Strange and Jenkins, 2013) who propose that languages (e.g., English, German) with more crowded vowel spaces rely more on dynamic spectral patterns to maintain contrasts.

Beyond the first two formants, which all of the aforementioned research has emphasized as major acoustic correlates of vowel identification, the third formant (F3) and vowel duration have been reported to be additional cues used in vowel identification (e.g., Hillenbrand et al., 1995; 2001; Watson and Harrington, 1999). For example, Hillenbrand et al. (1995), who collected their data from /hVd/ syllables,

noted that the inclusion of vowel duration increased the separation accuracy of vowels by 12% in some cases; F3 appeared to have an influence, but not more than the inclusion of vowel duration.

Within research on Arabic, only one study (e.g., Al-Tamimi, 2007b) has so far been carried out on vowel dynamics, but its aim was not to research intrinsic dynamic cues; rather, it was focused on looking at extrinsic dynamic vowel variation in both production and perception (see also, Al-Tamimi, 2007a). Briefly, Al-Tamimi (2007b) investigated the role of static cues compared to dynamic cues (e.g., formant slopes) in the classification of vowel systems in Jordanian Arabic (JA) and Moroccan Arabic (MA) dialects. The aim was to determine whether the dynamic nature of the CV transition is an important cue for vowel identification/discrimination. The discriminant analysis results revealed that while static cues permitted the discrimination of vowels in both dialects (74.85% for JA and 80.4% for MA), dynamic cues improved the classification accuracy by 13% for JA and 5% for MA. Hence, this previous study constitutes the first step into the field of intrinsic dynamic cues in the Arabic language, and the current study aims to investigate the Hijazi Arabic (HA) vowel system, which has not been studied acoustically before.

II. THE CURRENT STUDY

Arabic dialects vary in the size and makeup of their vowel system (Newman and Verhoeven, 2002). For instance, Syrian Arabic has 11 vowels: /u:, i:, e:, a, o, a:, ə, i, u, o, e/ (Almbark and Hellmuth, 2015), whereas MA includes only five vowels: /ʊ, u:, i:, a:, ə/ (Al-Tamimi, 2007a,b). Therefore, it is necessary to examine the characteristics of Arabic varieties separately when exploring the role of static and dynamic cues in their identification. HA is considered one of the main spoken dialects in the Kingdom of Saudi Arabia. It is spoken in the northwest of Saudi Arabia in various cities, such as Taif, Jeddah, Medina, and Makkah (Alzaidi, 2014; Abdoh, 2011). HA has eight vowels—namely /i:, a:, u:, i, a, u, e:, o:/ (Jarrah, 1993; Mousa, 1994; Abdoh, 2011; see Figure 1 below).

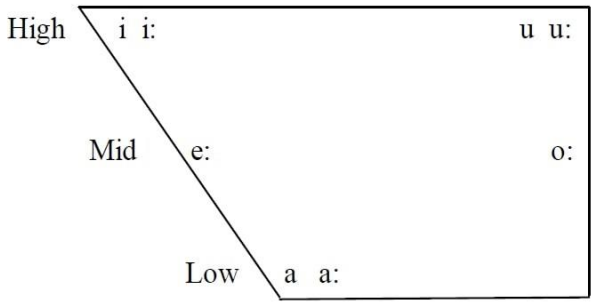


FIG. 1. Hijazi Arabic vowel system (adapted from Abdoh, 2011).

While early and impressionistic studies of Arabic mostly focused on phonological length differences between long and short Arabic vowels, more and more experimental research is demonstrating an added tense–lax contrast, showing that Arabic vowels differ in both quantity and quality (e.g., Al-Tamimi, 2007a,b; Almbark and Hellmuth, 2015). This is not surprising when considering the fact that articulatory duration and effort are often interlinked. For instance, greater articulatory effort in tense vowel production typically manifest as greater distinction in quality and longer duration (Chomsky and Halle, 1968). In other words, tense sound production requires the articulatory organs to maintain a given configuration for longer compared to non-tense sounds. A difference in quality between long and short Arabic vowels may also be a by-product of a difference in the articulatory effort required for long vowels.

The purpose of the current study is to investigate to what extent the static and dynamic cues improve the classification of HA vowels as well as to what extent vowel duration and F3 act as additional cues to classification accuracy. A further purpose is to look at the quality of HA vowels to investigate if there is a difference between them in term of quality as well as quantity.

III. METHODOLOGY

A. Subjects and material

The participants were 12 male native HA speakers, aged 18 to 30, who were born and raised in Hijaz, Saudi Arabia. They reported no history of speech and/or language disorders. Recordings were made on a Zoom Digital H1 Handy Recorder with a sampling rate of 44,100 Hz and 16-bit amplitude resolution. The subjects were placed in a soundproof room at Taibah University. The HA speakers were asked to produce all vowels in a monosyllabic /hVd/ context within the phrase /ktoːb _____ marteːn/, which means "Write ___ twice" (see Table I). Together, the HA stimuli comprised 5 repetitions × 8 vowels × 12 HA male participants = 480 items. It was difficult to put all of the HA vowels into real /hVd/ words in HA; therefore, the nearest real HA words that have the same target vowels, such as /xoːd/ and /zeːd/, were used.

TABLE I. The set of target words presented to the participants.

HA vowel	Target word	HA Arabic presentation	English gloss
/u:/	/hu:d/	هود	Male name
/i:/	/hi:d/	هيد	Calm down
/e:/	/ze:d/	زيد	Male name
/o:/	/xo:d/	خود	Take
/a:/	/ha:d/	هاد	Relaxed
/i/	/hidd/	هِد	Destroy!
/a/	/hadd/	هَد	Drive slowly
/u/	/hudd/	هُد	To hit someone's head

B. Acoustic analyses

An acoustic analysis was done using PRAAT (Boersma and Weenink, 1992-2019). All formant tracks were obtained using a 0.025s window length, 50 Hz pre-emphasis, and 5000-Hz maximum formant frequency. The "burg" method was used to extract formant frequencies, with a maximum formant number

of 5 (i.e., obtaining 5 poles within 5000 Hz). For the purposes of this research, the vowel duration and the first three formant values were automatically extracted with the aid of a PRAAT script specifically created for this study by the first and second authors. The onset and offset of the vocalic segment were manually labeled for each /hVd/ syllable. Vowel onsets were labeled at the end of the noise for /h/, before the first positive peak in the periodic waveform. The offsets of the vowels were set as the end of the periodicity in the waveform before the stop closure of /d/. The vowel duration between the start and end boundaries was measured as the duration (in ms). Vowel segmentations included the entirety of the vowel, so transitions were also included in the segmentation and duration measures, but not in the formant analyses reported below. This was done to ensure that the measurements taken along the vowel trajectory were all at the same time points within the vowels. For example, if the vowels were segmented without the consonant transitions, measurements could potentially begin at different time points within the vowel. This would remove some of the advantages of using the proportional distance approach. Therefore, we measured the entirety of the vowel (0% to 100%), but measurements of 0%–19% and 81%–100% were not included when reporting on formant measurements (e.g., Cardoso, 2015). F1, F2, and F3 were extracted from one location (50% for the static model), two locations (20% and 80% for the offset, direction, and slope models), and three locations (20%, 50%, and 80% for the three-point model) across the vowel duration.

To investigate the amount of spectral shifts for HA vowels in the offset model, the first three formants were computed as

$$(1)\ \sqrt{(\text{Offset80\% - Onset20\%})^2}$$

For the direction model, the first three formants were computed as

$$(2)\ (\text{Offset80\% - Onset20\%}),$$

whereas for the slope model, the first three formants were computed as

$$(3)\ (\text{Offset80\% - Onset20\%})/\text{duration}$$

All formant values were checked manually to ensure the accuracy of the results, and any errors in formant estimation were corrected by hand. For example, in some cases, PRAAT reads close values of F2 as F1 and also reads the F3 values as F2. To mitigate PRAAT's measurement errors, all formant frequencies were visually verified for errors in extraction, and in cases where formants were misidentified by the automatic procedure, they were manually corrected.

C. Statistical analyses

Two types of statistical techniques were used to evaluate the differences in the data—namely, linear mixed-effects modeling (LMM), which was then followed by discriminant analyses as a classification tool. All figures and analyses were created and run in RStudio (version 1.2.1335; 2019) and R Core Team (version 3.6.1; 2019) with packages ggplot2 (version 3.2.1; Wickham, 2016), dplyr (version 0.8.3; Wickham et al., 2019), lme4 (version 1.1.21; Bates et al., 2015), emmeans (version 1.3.5.1; Length, 2019), and MASS (version 7.3.51; Venables and Ripley, 2002). LMM was run using the package lme4 (Bates et al., 2015). Our outcome was each of the 12 acoustic correlates (F1, F2, and F3 for the static and for each of the three dynamic cues). Our fixed effect was the vowel identity (with eight levels). Our random effect was the subject. For each acoustic correlate, we ran three versions: a null (or an average) model, an intercept model with the fixed effect, and, finally, a slope model with the fixed effect and by-subject adjustment for the fixed effect, the vowel. Through a log likelihood model comparison, it was apparent that in all cases, the intercept model improved the model fit compared to the null model and that the slope model did not improve the model fit. This is likely because we used an /hVd/ environment that did not have much of an effect on vowel production and speakers did not vary in how they produced the various patterns observed below, despite, of course, the presence of idiosyncrasies, which are taken into account by our models. All of our results are based on an intercept-only model with the following specification: lmer(outcome ~ vowel + (1|Subject), data = data). Following our LMMs, we used the package emmeans (Length, 2019) to report on the pairwise comparisons with the false discovery rate adjustments for multiple comparisons. These

post-hoc tests are based on our LMM model, with estimated marginal means and standard error (SE). In the results section, we report on the model comparison followed by the pairwise comparison results.

The next step was applying the discriminant analyses as a classification tool. We used the function qda from the package MASS (Venables and Ripley, 2002) to obtain the quadratic discriminant analyses with a *leave-one-out* cross-validation, or "jackknife" (Hillenbrand et al., 1995). Discriminant analyses evaluate the robustness in the observed differences between vowels by looking at the combination of predictors used. The analysis performs a multivariate analysis of variance on the combination of predictors and creates discriminant functions that are used to separate the vowels. These discriminant functions can be either positively or negatively correlated with each of the predictors. Then the discriminant analysis tries to separate the vowels into multiple groupings to arrive to an optimal separation between the categories. Cross-validation was performed at the prediction stage and is a way to evaluate the classification accuracy as if done on unseen data. For each of the models below, we used the vowels as categories to be classified and each of the formant frequencies or each of the formulae and vowel duration outputs as predictors. For example, the models presented in TABLE II below used the full eight vowels as categories and the following predictors as input to each of the discriminant analyses: For the static model, we entered the formant values sampled from vowel midpoint at 50%; for the direction and slope models, we entered the results of their formulae above; for the offset model, we entered the formant values sampled from vowel onset (at 20%) and offset (at 80%); and finally, for the three-point model, we entered the formant values sampled from vowel onset (at 20%), midpoint (at 50%) , and offset (at 80%). In all dynamic measures, we compared the models using F1+F2 with those that used F1+F2+F3 (with or without the duration). For the tense versus lax results (see Table III), we used only six vowels as categories. For vowel pairs, two vowels were used in each sub-model (see Table IV).

IV. RESULTS

A. Overall patterns

This section presents the descriptive results of the static and dynamic cues. A full summary of the results for duration and the first three formant values of HA vowels can be found in the Appendix.

1. Static cues

Beginning with the static model, the results of the model comparison showed a clear improvement to the model fit when using the vowel as a fixed effect (F1: $\chi^2(7)=1083.4$, $p<0.0001$; F2: $\chi^2(7)=1604.1$, $p<0.0001$; F3: $\chi^2(7)=87.7$, $p<0.0001$). Figure 2 displays a scatterplot of the first two formant values for all of the HA vowels across all of the subjects. It shows a clear and significant separation in the vowel space between HA vowels, in particular the short and long pairs. The results of the pairwise comparisons for the /a/ and /a:/ pair showed an overall lower F1 and higher F2 frequencies for /a/ (for F1, there was a difference of -67.2 Hz (SE = 6), $p<0.0001$, and for F2, a difference of 264.2 Hz (SE = 16.2), $p<0.0001$). For the /i/ and /i:/ pair, the results showed an overall higher F1 and lower F2 frequencies for /i/ (F1 had a difference of 78.3 Hz (SE = 6), $p<0.0001$, and F2 had a difference of -272.4 Hz (SE = 16.2), $p<0.0001$). For the pair /u/ and /u:/, the results showed an overall higher F1 and higher F2 frequencies for /u/ (for F1, there was a difference of 73.4 Hz (SE = 6), $p<0.0001$, and for F2, a difference of 336.2 Hz (SE = 16.2), $p<0.0001$). These results showed a clear difference between the short and long vowels in terms of quality, with /i a u/ proving to be centralized compared with their long counterparts, potentially suggesting a lax quality.

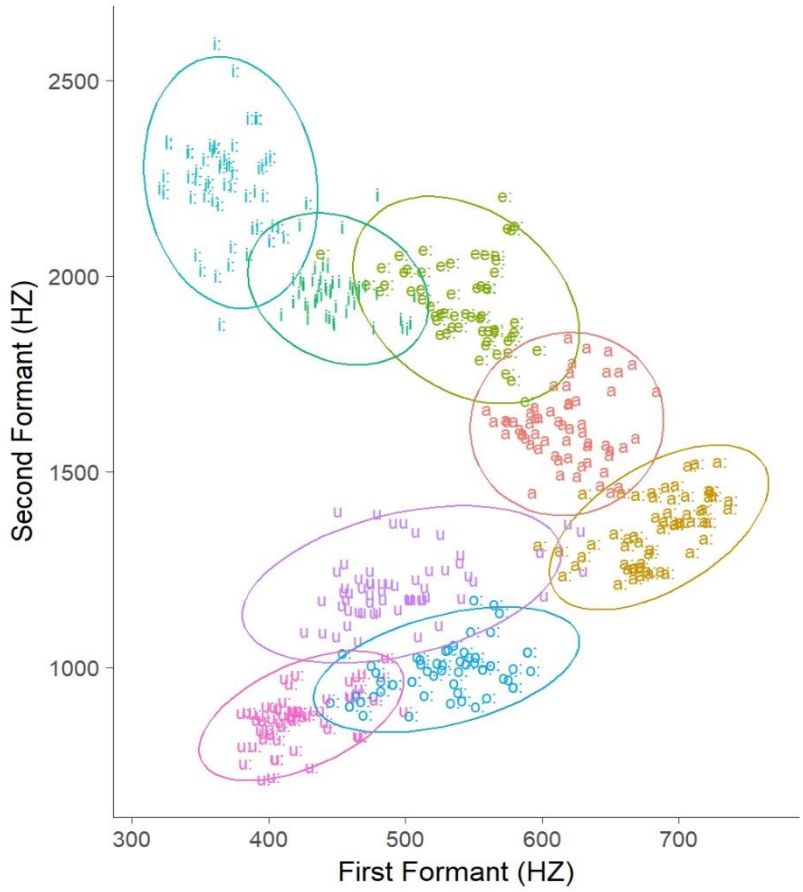


FIG. 2. Scatterplot of the midpoints of the first two formant values of Hijazi Arabic vowels.

2. Dynamic cues

Moving on to the dynamic models and starting with the offset model, the results of the model comparison showed a clear improvement to the model fit when using the vowel as a fixed effect (F1: $\chi^2(7)=423.3$, $p<0.0001$; F2: $\chi^2(7)=427.1$, $p<0.0001$; F3: $\chi^2(7)=23.3$, $p<0.002$). As can be seen from Figure 3, the degree of overall spectral change is important and is up to 600 Hz for F2, up to 200 Hz for F1, and up to 400 Hz for F3. When looking at vowel pairs, the results of the pairwise comparisons showed that for some comparisons, the differences were statistically significant. For F1, only /a/ versus /a:/ showed a statistically significant difference, with /a/ having a higher positive difference by 70.2 (SE = 4.5), $p<0.0001$. For F2, the pairs /i/ versus /i:/ and /u/ versus /u:/ showed a statistically significant difference for offset values: /i/ showed a negative difference of -36.8 (SE = 4.6), $p=0.035$, compared to /i:/; /u/ showed a positive difference of 147.7 (SE = 4.6), $p<0.0001$, compared to /u:/. For F3, there were no statistical differences

between the vowel pairs. As was found for the static results, these patterns indicate a difference between specific vowel pairs and formant frequencies that is possibly related to a tense–lax distinction, alongside the durational contrast, but is not robust for all vowel pairs or across all formant frequencies.

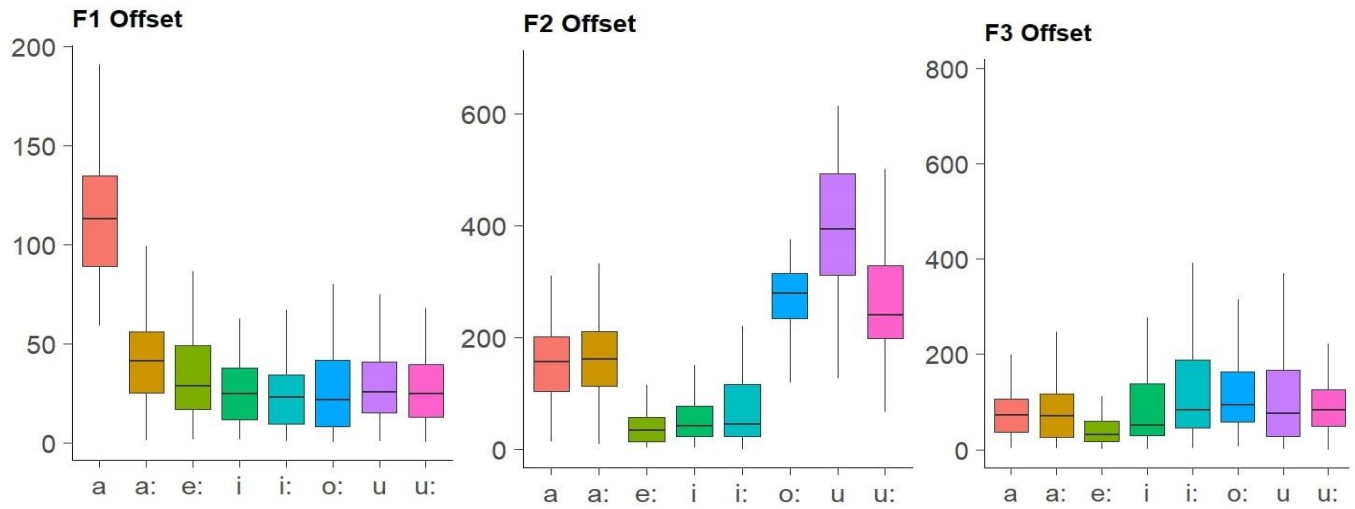
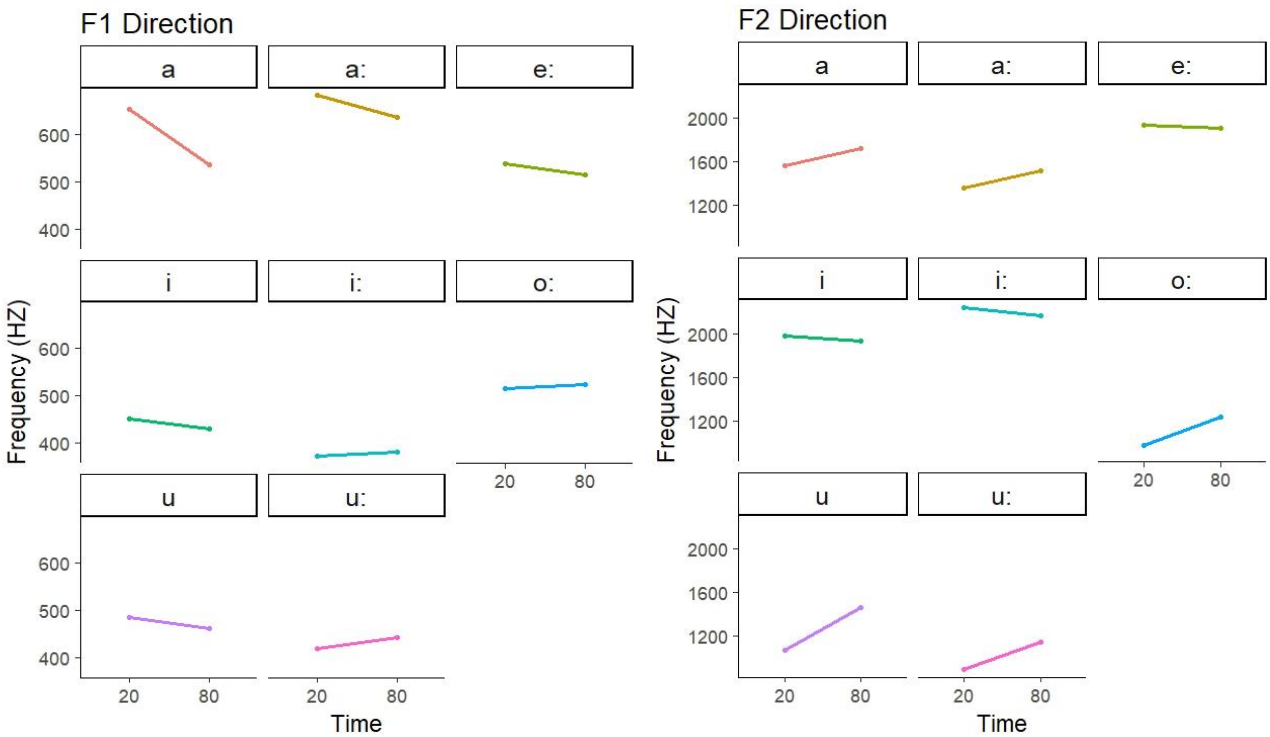


FIG. 3. Boxplot of the offset model for the eight Hijazi Arabic vowels.

Regarding the results of the direction model, Figure 4 presents the direction of the spectral change, which shows variation among HA vowels, with each vowel having its own dynamic feature for each formant (thinner lines are for individual tokens and the thicker lines are the means). The results of the model comparison showed a clear improvement to the model fit when using the vowel as a fixed effect (F1: $\chi^2(7)=492.8$, $p<0.0001$; F2: $\chi^2(7)=554.8$, $p<0.0001$; F3: $\chi^2(7)=40.9$, $p<0.0001$). Most importantly, the short vowels' F1 directions displayed a significantly decreasing spectral shift compared to their long counterparts. In addition, both /i/ and /u/ had falling transitions compared with /i:/ and /u:/; /a/ and /a:/ had similar falling transitions, but /a/ had a steeper falling transition. Using pairwise comparisons on vowel pairs showed that for the pair /a/ and /a:/, there was an overall higher difference related to the steeper transition of /a/ only for F1, with no differences for F2 or F3 (for F1, the difference was 70.8 Hz (SE = 5.6), $p<0.0001$). For the pair /i/ and /i:/, the results showed an overall higher direction value for F1 and a lower one for F2 and F3 for /i/ (for F1, the difference was 30.5 Hz (SE = 5.6), $p<0.0001$; for F2, the difference was -37.4 Hz (SE = 19.7), $p=0.065$; and for F3, the difference was -91 Hz (SE = 6.5), $p<0.003$). For the

pair /u/ and /u:/, the results showed an overall higher direction value for F1 and lower for F2 for /u/ (for F1, the difference was 46.7 Hz (SE = 5.6), $p<0.0001$, and for F2, it was -133.2 Hz (SE = 19.7), $p<0.0001$). However, F2 direction showed that low and back vowels had increasing slopes, whereas front ones had a decreasing slope while the F3 direction changes were not systematic.



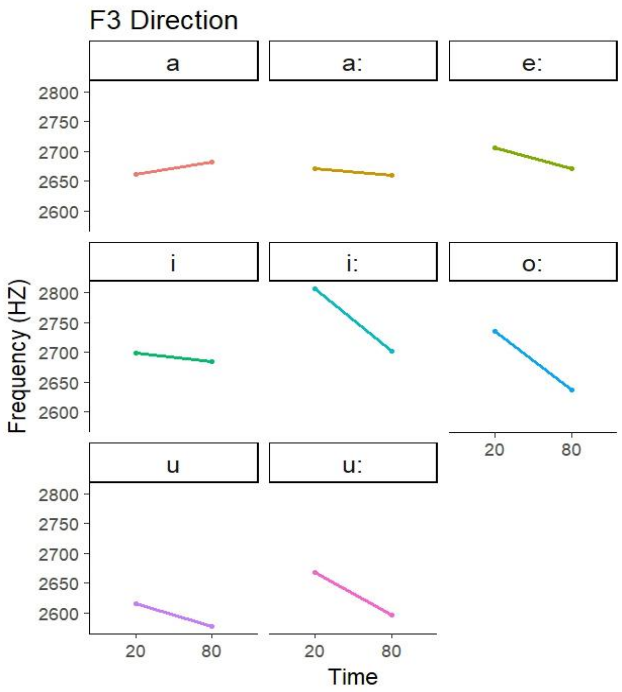


FIG. 4. Results of the direction model for the eight Hijazi Arabic vowels.

The same can be seen with respect to the slope model. This model showed significant variation across the vowels (see Figure 5), and each HA vowel had a unique dynamic feature in terms of the slope for each formant. The results of the model comparison showed a clear improvement to the model fit when using the vowel as a fixed effect (F1: $\chi^2(7)=624.1$, $p<0.0001$; F2: $\chi^2(7)=561.3$, $p<0.0001$; F3: $\chi^2(7)=67.5$, $p<0.0001$). Looking into the results in more detail showed that the F1 slopes with the short vowels appeared to differ significantly from their long counterparts. The results observed here mirror those seen above in the direction model. Recall that the two formulae are similar with the only difference being that the slope model uses the output from the direction model and divides it by the duration. The results of the pairwise comparisons on the vowel pairs showed that for the pair /a/ and /a:/, there was an overall higher difference that is related to the steeper transition of /a/ only for F1 with no differences for F2 or F3 (the F1 had a difference of 0.95 (SE = 0.05), $p<0.0001$). For the pair /i/ and /i:/, the results showed an overall higher slope value for F1 and lower ones for F2 and F3 for /i/ (for F1, there was a difference of 0.29 (SE = 0.05), $p<0.0001$; for F2, a difference of -0.12 Hz (SE =0.04), $p<0.006$; and for F3, a difference of -0.04 Hz (SE = 0.02), $p=0.047$). For the pair /u/ and /u:/, the results showed an overall higher F1 and lower F2 slope for /u/ (for F1, there was a difference of 0.39 Hz (SE = 0.05), $p<0.0001$, and for F2, a difference of -0.21 Hz (SE = 0.4), $p<0.0001$). Similar to the F2 direction, the F2 of the front vowels had a falling slope, unlike the low and back vowels, which had raising slopes. The F3 slope exhibited few changes which were not systematic.

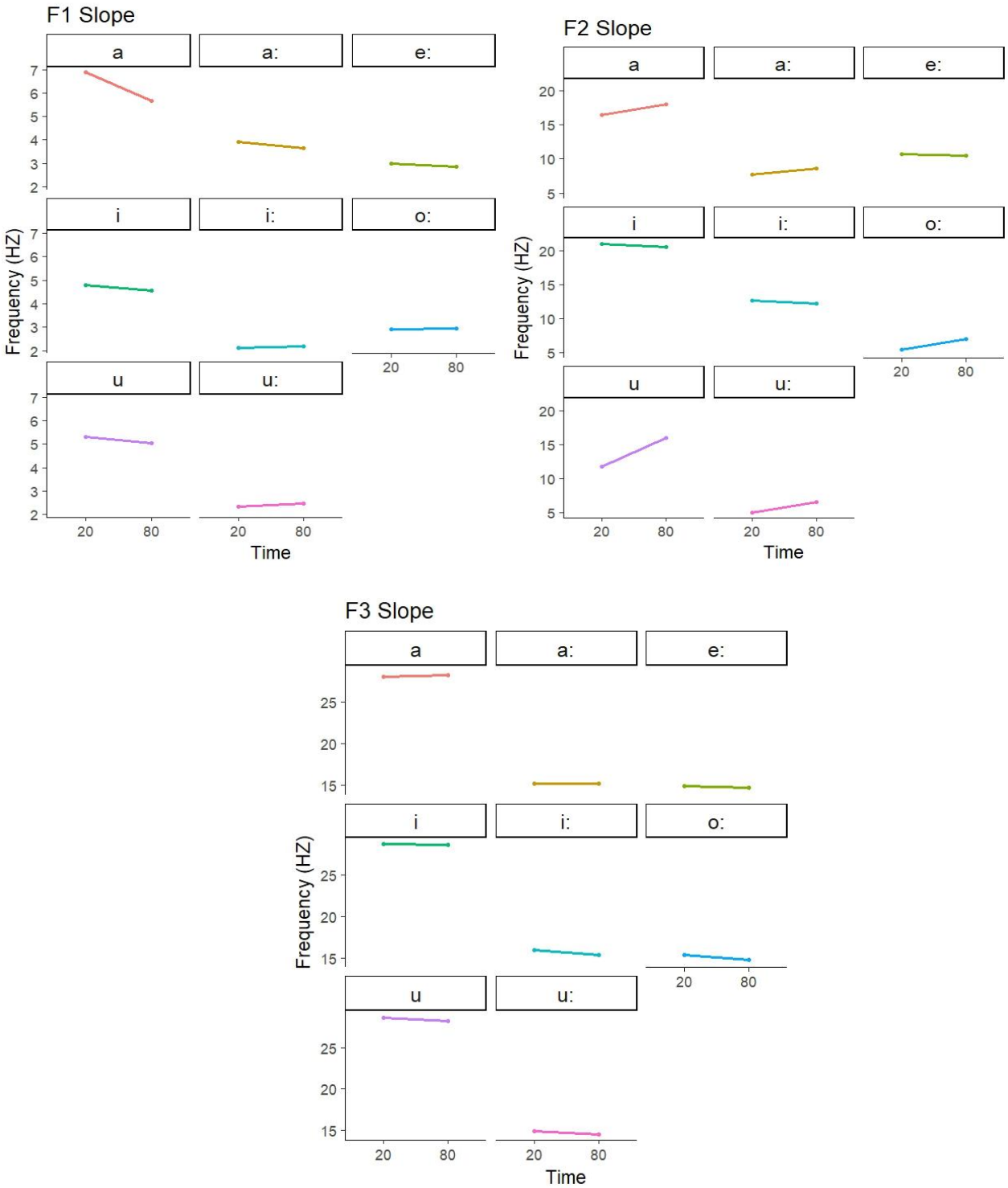


FIG. 5. Results of the slope model for the eight Hijazi Arabic vowels.

B. Discriminant analysis

The results of the three approaches, alongside the static approach, were evaluated via discriminant function analyses as a way to evaluate the degree of separation between HA vowels and to quantify the

classification rates using each of the models. We ran discriminant analyses in three stages: We started by evaluating the discrimination between all eight HA vowels, then we evaluated the discrimination between the lax and tense vowels as a group (e.g., group 1: /i a u/ versus group 2: /i: a: u:/) and, finally, between HA vowel pairs in three groups (group 1: /i:/ versus /i/, group 2: /u:/ versus /u/, and group 3: /a:/ versus /a/).

Starting with the first model, the discriminant analysis results showed that using the three point approach (e.g., 20-50-80%) with F1, F2, and F3 (with and without the duration) resulted in the highest classification accuracy (from 93% to 97%) for all eight HA vowels, followed by the offset approach (from 91% to 97%), then the static approach (from 90% to 96%), and, finally, the other VISC approaches—namely, the slope approach (from 61% to 74%) and the direction approach (from 57% to 74%) (see Table II).

TABLE II. Discriminant analysis results showing the classification accuracy of vowels trained on various combinations of parameters for model 1 ("No Dur" indicates that the duration was not included, whereas "Dur" means the duration was included).

	Static		**Direction**		**Slope**		**Offset**		**Three-Point model**	
	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur
F1 F2	90	96	57	73	61	74	91	97	93	97
F1 F2 F3	92	96	59	74	61	74	92	97	95	97

Moving on to model 2, where we aimed to discriminate between lax and tense vowels as two groups, the results showed a higher improvement in the classification accuracy in comparison to Table II. The three-point approach and the static approach achieved the best rates (97–99% and 96–99%, respectively), followed by the offset approach (between 94% and 98%) in most cases when the combination of F1, F2, and F3 and vowel duration were used. On the other hand, the classification accuracy rate for the slope

approach was between 77% and 91% whereas it was between 73% and 90% for the direction approach (see Table III).

TABLE III. Classification rates of Hijazi Arabic vowels for model 2 (lax versus tense).

	Static		**Direction**		**Slope**		**Offset**		**Three-Point model**	
	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur
F1 F2	96	98	73	90	77	91	94	98	97	99
F1 F2 F3	97	99	73	90	77	91	95	98	98	99

Moving on to the third model, where we looked at the vowel pairs (e.g., /i/ vs. /i:/, /u/ vs. /u:/, and /a/ vs. /a:/), the results showed a noticeable improvement in the classification accuracy compared to Table II and III. The three-point approach had a better rate (99%) in most cases when the vowel duration and the combination of the first three formants were used, followed by the static approach (between 96% and 99%), then by the offset approach (between 95% and 99%). For other VISC approaches, the average rate of discrimination between the HA vowel pairs was between 78% and 99% for the slope approach, whereas it was between 74% and 99% for the direction approach (see Table IV).

Table IV: The correct classification rates of Hijazi Arabic vowel pairs /i/ vs. /i:/, /u/ vs. /u:/, and /a/ vs. /a:/.

		Static		**Direction**		**Slope**		**Offset**		**Three-Point model**	
		No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur	No Dur	Dur
/i/ vs. /i:/	**F1 F2**	99	99	86	99	86	99	98	99	99	99
	F1 F2 F3	99	99	90	99	89	99	99	99	99	99
/u/	**F1 F2**	96	99	74	97	78	98	96	99	99	99

vs. /u:/	**F1 F2 F3**	96	99	75	97	78	98	96	99	99	99
/a/	**F1 F2**	97	99	82	97	95	98	95	99	99	99
vs. /a:/	**F1 F2 F3**	98	99	84	97	95	98	99	99	99	99

Looking at the results of the addition (or absence) of vowel duration in each of the three tables above shows that vowel duration played an important role in classification accuracy of all eight HA vowels, and its inclusion with the formant frequencies in any model led to a substantial improvement in vowel separation by up to 15% (average +8%). On the other hand, the role of F3 appeared to have little influence on the classification accuracy of HA vowels, with its inclusion in some models improving the classification rates of HA vowels by between 1% and 2% overall (average about 1%). The average correct classification rates for the proposed approaches for all eight HA vowels (See Table II) are as follows: The three-point approach yielded the highest classification accuracy (average 95.5%), followed by the offset approach (average 94.25%), then the static approach (average 93.5%), and then the other VISC approaches—namely, the slope approach (average 67.5%) and direction approach (average 65.75%).

V. DISCUSSION AND CONCLUSION

The main purpose of this paper was to evaluate the importance of static and dynamic cues in Arabic vowels, exploring the role of VISC approaches and the three-point approach in the classification of HA vowels alongside vowel duration and F3. A further purpose was to explore if there is a tense–lax contrast in HA vowels alongside a phonological length contrast.

The data demonstrate that the three-point approach is the best approach and is the most accurate for classifying HA vowels in all three models (with an average classification accuracy of 98.1%) in comparison

to the other approaches, namely, the static and other VISC approaches. Such a finding provides support for the three-point approach and is in line with many previous studies (e.g., Huang, 1992; Zahorian and Jagharghi, 1993; Harrington and Cassidy, 1994; Hillenbrand et al., 1995; Ferguson and Kewley-Port, 2002; Yuan, 2013) that concluded that monophthong vowels of different quality can obtain better identification when their acoustic parameters are taken from three points (onset + midpoint + offset). The offset model, on the other hand, comes in second as the best approach for obtaining better identification of HA vowels (with an average classification rate of 97.5%). It supports previous research (e.g., Hillenbrand et al., 1995; Hillenbrand and Nearey, 1999; Hillenbrand et al., 2001) and uses a couple of locations, one early and one late in the syllable (onset + offset), leading to higher correct classification rates than using one point (midpoint).

Interestingly though, the data reveal that the static approach was sufficient and obtained higher classification accuracies (with average of 97.1%) for classifying HA vowels than the other VISC models based on the direction and slope approaches. Such a result is contrary to the expectations of other studies (e.g., Nearey and Assmann, 1986; Arnaud et al., 2011) that reported better identification of vowels in direction and slope approaches and incorporated the spectral change of the vowel rather than a measurement sampled at a single time. The interpretation of this result could be illustrated as follows: Those studies that found that direction and slope approaches outperformed the single-point approaches in classification accuracy examined both models in different phonetic environments than /hVd/, and according to Elvin et al. (2016), the /hVd/ context is acoustically least comparable to other consonantal contexts. Elvin et al. (2016) found that by using the discriminant analysis, the recognition scores are least accurate from tokens taken from /hVd/ compared to other contexts. This could be due to the phonological voicing status of the following coda, which might significantly alter spectral characteristics and vowel duration. On the other hand, it is possible that /hVd/ may be a better predictor for other voiced coda contexts. Similarly, there are other studies (e.g., Harrington and Cassidy, 1994; Watson and Harrington, 1999) that found that spectral

information from the midpoint was sufficient for monophthong identification using the /hVd/ context. Together, these findings suggest that experimental results on vowels with other consonantal context transitions, which provide additional information regarding the vowel's phonetic identity, are identified more accurately by all VISC models than vowels in isolation or /hVd/ (Oh, 2013), as /hVd/ syllables do not contain many spectral changes. The consonantal environments are known to affect the vowel formant values (Hillenbrand et al., 2001). For example, a study by Stevens and House (1963) showed that vowel formant patterns in isolation are the same as in /hVd/, which shows a negligible effect on vowels, whereas formant values exhibit spectral changes when they are in the environment of a more comprehensive list of consonants. Hence, it is likely that the differences in findings between this paper and other studies' findings (e.g., Nearey and Assmann, 1986; Arnaud et al., 2011) are due to contextual differences, and certainly more studies are needed with a great number of voiced coda environments to determine the nature of such acoustic differences.

The slope and direction models provide some insight into and a better overview of the characterization of dynamic cues of the HA vowels and how each vowel has its own dynamic feature, particularly the tense/lax vowels. This result is consistent with Watson and Harrington (1999) and Slifka (2003), who found that using formant trajectory was useful for the within-class separation of lax/tense vowels. Similarly, this study found that F1 slope and direction of the HA short vowels are significantly different from their long vowel counterparts. These results support other studies (e.g., Al-Tamimi, 2007a,b; Almbark and Hellmuth, 2015) that argue that Arabic short and long vowels are different in terms of both their quality and quantity. This study found that HA vowels displayed great spectral movement, as was found in other studies that have noted that speech dynamics are greater for speakers with a sparse vowel system (e.g., Manuel, 1990; Meunier et al., 2003; Al-Tamimi and Ferragne, 2005; Jin and Liu, 2013). This may be due to low-density languages having more space and freedom to produce their vowels compared to high-density languages.

Although the effectiveness of the first two formant frequencies in vowel identification is indisputable, this study highlights the fact that vowel duration is the most important additional cue for the classification accuracy of HA vowels. On the one hand, this conclusion should not be so surprising given previous studies that noted that including vowel duration increased the separation of the vowels when using a discriminant analysis (e.g., Hillenbrand et al., 1995; 2001; Watson and Harrington, 1999). However, vowel duration in this study has proven to have more influence than has been found elsewhere, with a substantial improvement in vowel separation (up to 15%), whereas in Hillenbrand et al.'s (1995) study, it was only up to 7.9%. This can be explained by considering the phonological role of vowel duration as a cue to distinguishing short and long vowels in HA vowels. Regarding the role of F3, it appears to have little influence on the classification accuracy of HA vowels, which is in agreement with other studies (e.g., Hillenbrand et al., 1995), and this may be due to the fact that F3 is a better index for lip rounding and speaker physiology than inherent vowel identity.

In sum, our results are found to be more consistent with dynamic theories of vowels, as they provide evidence that monophthong vowels are dynamic and that vowel duration is the most useful additional feature to differentiate between phonemes. As mentioned earlier, this research is the first step in looking at vowels as intrinsically dynamic in the Arabic language. This study's results could be extended to look at contexts beyond /hVd/, as suggested by many researchers (e.g., Hillenbrand et al., 1995; Watson and Harrington, 1999), in order to dig deeper into dynamic properties in various consonantal contexts and provide further comparative research, which will be our next step.

ACKNOWLEDGMENTS

This work was supported by a scholarship from the Saudi Cultural Bureau to the first author (WA) and from the Leverhulme International Academic Fellowship (IAF-2018-016) to the second author (JA).

We thank two anonymous reviewers, Dr. Richard A. Wright and Dr. Lauren Ackerman for their insightful comments on a previous version of the paper. We thank all of our subjects who participated in this study.

APPENDIX

TABLE V. Average of the formant frequencies (at 20% onset, 50% midpoint, and 80% offset) and vowel duration for each Hijazi Arabic vowel.

		F1 (Hz)	**F2 (Hz)**	**F3 (Hz)**	**Duration (ms)**
/u:/	Onset	418.72	892.76	2668.02	181.16
	Mid	423.88	877.27	2725.57	
	Offset	442.60	1148.92	2596.93	
/i:/	Onset	373.18	2239.11	2807.11	180.23
	Mid	372.29	2240.24	2775.16	
	Offset	382.35	2162.79	2702.25	
/e:/	Onset	538.53	1938.73	2705.82	185.13
	Mid	544.18	1939.69	2703.97	
	Offset	515.05	1905.52	2670.73	
/o:/	Onset	516.70	977.06	2735.93	182.27
	Mid	529.73	994.77	2755.69	
	Offset	523.61	1237.93	2636.83	
/a:/	Onset	682.21	1357.19	2670.70	178.28
	Mid	685.08	1358.68	2667.81	
	Offset	636.10	1514.81	2660.22	
/i/	Onset	450.81	1972.94	2698.86	96.25
	Mid	450.60	1967.88	2654.97	
	Offset	429.48	1934.06	2685.00	

/a/	Onset	652.25	1566.02	2661.12	96.77
	Mid	617.91	1622.90	2639.97	
	Offset	535.34	1717.53	2682.04	
/u/	Onset	485.42	1075.68	2615.28	93.98
	Mid	497.28	1213.44	2559.26	
	Offset	462.57	1465.09	2576.95	

Abdoh, A. (**2011**). "A study of the phonological structure and representation of first words in Arabic," Ph.D. dissertation, University of Leicester, Leicester.

Almbark, R., and Hellmuth, S. (**2015**). "Acoustic analysis of the Syrian vowel system," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, University of Glasgow. ISBN 9780852619414.

Al-Tamimi, J. (**2007a**). "*Indices dynamiques et perception des voyelles: étude translinguistique en arabe dialectal et en français* ("Dynamic indices and vowel perception: translinguistic study in Arabic and in French dialects,") Unpublished PhD dissertation, University Lyon, 2 (accessible here in French :http://theses.univ-lyon2.fr/documents/lyon2/2007/al-tamimi_je).

Al-Tamimi, J. (**2007b**). "Static and Dynamic cues in Vowel Production: a cross dialectal study in Jordanian and Moroccan Arabic," in *proceedings of the 16th ICPHS*, Saarbrücken, Germany, 541-544.

Al-Tamimi, J. and Ferragne, E., (**2005**). "Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French," in *proceedings 9th EuroSpeech*, Lisbon, pp. 2465-2468.

Alzaidi, M. (**2014**). "Information Structure and Intonation in Hijazi Arabic," Ph.D. dissertation, University of Essex, Colchester.

Arnaud, V., Sigouin, C., and Roy, J. P. (**2011**). "Acoustic description of Quebec French high vowels: First results," in *Proceedings of the 17th ICPhS*, Hong Kong, 244-247.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4,". J Stat Softw. **67**(1), 1–48. doi: 10.18637/jss.v067.i01. R package version 1.1.21. https://CRAN.R-project.org/package=lme4

Boersma, P., and Weenink, D. (**1992–2019**). Praat: Doing phonetics by computer.

Cardoso, A., B. (**2015**). "Dialectology, phonology, diachrony: Liverpool English realisations of PRICE and MOUTH," Ph.D. dissertation, University of Essex. Colchester.

Chomsky, N., and Halle, M. (**1968**). *The sound pattern of English*. London: Harper & Row.

Elvin, J., Williams, D., and Escudero, P. (**2016**). "Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English," J. Acoust. Soc. Am. **140**(1), 576-581.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**(1), 259-271.

Gottfried, M., Miller, J. D., and Meyer, D. J. (**1993**). "Three approaches to the classification of American English vowels," J. Phone. **21**: 205-229.

H, Wickham. (**2016**). ggplot2: *Elegant Graphics for Data Analysis*. Springer-Verlag New York. R package version 3.2.1. https://ggplot2.tidyverse.org

Harrington, J., and Cassidy, S. (**1994**). "Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English," Lang. Speech. **37**(4), 357-373.

Hillenbrand, J. M. (**2013**). "Static and dynamic approaches to vowel perception," in *Vowel Inherent Spectral Change* edited by G. S. Morrison and P. F. Assmann (Springer, Berlin), pp. 9–30.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hvd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509-3523.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**(2), 748-763.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**: 3099-3111.

Huang, C. B. (**1992**). "Modelling human vowel identification using aspects of formant trajectory and context," in *Speech perception, production and linguistic structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (IOS, Amsterdam), pp. 43–61.

Jarrah, M. A. (**1993**). "The Phonology of Madina Hijazi Arabic: A Non-linear Analysis," Ph.D. dissertation, University of Essex, Colchester.

Jin, S. H., and Liu, C. (**2013**). "The vowel inherent spectral change of English vowels spoken by native and non-native speakers," J. Acoust. Soc. Am. **97**: 3099-3111.**133**(5), EL363-EL369.

Manuel, S. Y. (**1990**). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages," J. Acoust. Soc. Am. **88**(3), 1286-1298.

Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. (**2003**). "Production and perception of vowels: does the density of the system play a role?," in *Proceedings of the 15th ICPhS*. Barcelona, 723-726.

Morrison, G. S., and Nearey, T. M. (**2007**). "Testing theories of vowel inherent spectral change," J. Acoust. Soc. Am. **122** (1): EL 15-22.

Morrison, S., and Assmann, P. (**2012**). *Vowel inherent spectral change* (Springer Science & Business Media).

Mousa, A. (**1994**). "The Interphonolgy of Saudi Learners of English," Ph.D. dissertation, University of Essex, Colchester.

Munro, M. J. (**1993**). "Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings," Lang. Speech. **36**(1), 39-66.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**(5), 1297-1308

Newman, D., and Verhoeven, J. (**2002**). "Frequency analysis of Arabic vowels in connected speech," Antwerp Papers in Linguistics. 100: 77-86.

Oh, E. (**2013**). "Dynamic spectral patterns of American English front monophthong vowels produced by Korean-English bilingual speakers and Korean late learners of English," Linguistic Research, **30**(2), 293-312.

Peterson, G. E., and Barney, H. L. **(1952)**. "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175-184.

R Core Team., (**2019**). R: A language and environment for statistical computing (version 3.6.1). Vienna, Austria: R Foundation for Statistical Computing. [Software Resource]. ISBN 3-900051-07-0. https://www.R-project.org/

RStudio., (**2019**). Rstudio: Integrated development environment for R (version 1.2.1335). Boston, MA. [Software Resource]. https://rstudio.com/

Length, R. (**2019**). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.5.1. https://CRAN.R-project.org/package=emmeans

Slifka, J. (**2003**). "Tense/lax vowel classification using dynamic spectral cues," in *Proceedings of the 15th ICPhS*. Barcelona, 921-924.

Stevens, K. N., and House, A. S. (**1963**) "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**: 111–128.

Strange, W. (**1989**). "Evolving theories of vowel perception," J. Acoust. Soc. Am. **85**(5), 2081-2087.

Strange, W., and Jenkins, J. (**2013**). "Dynamic specification of coarticulated vowels: Research chronology, theory, and hypotheses," in *Vowel Inherent Spectral Change* edited by G. S. Morrison and P. F. Assmann (Springer, Berlin), pp. 87-115.

Venables, N., and Ripley, D. (**2002**). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0. R package version 7.3.51https://CRAN.R-project.org/package=MASS

Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**: 458-468.

Wickham, H., François, R., Henry, L., and Müller, K. (**2019**). Dplyr: A grammar of data manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

Yuan, J. (**2013**). "The spectral dynamics of vowels in Mandarin Chinese," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, pp. 1193-1197.

Zahorian, S. A., and Jagharghi, A. J. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**(4), 1966-1982.